GAMMA-RAY LARGE AREA SPACE TELESCOPE
(GLAST)

SCIENCE ANALYSIS TOOLS
DATABASES REQUIREMENTS

JULY 26, 2002

Table of Contents

1 Purpose

This document describes the requirements for databases to be used with the science analysis tools for the GLAST LAT. These databases have been identified by the Science Tools Working Group (STWG). The analysis software will make queries of the databases, through specially devised access utilities. These databases will be populated by data generated from the LAT IOC and the GBM IOC.

2 Acronyms

API	Application Program Interface
DBMS	Database Management System
GBM	GLAST Burst Monitor
GIOC	GBM IOC
GRB	Gamma-Ray Burst
GSFC	Goddard Space Flight Center
GLAST	Gamma-ray Large Area Space Telescope
HEASARC	High Energy Astrophysics Science Archive Research Center
IOC	Instrument Operations Center
LAT	Large Area Telescope
LHEA	Laboratory for High Energy Astrophysics
LIOC	LAT IOC
NASA	National Aeronautics and Space Administration
SAA	South Atlantic Anomaly
SSC	Science Support Center
STWG	Science Tools Working Group
SWG	Science Working Group
TBD	To Be Determined
TBR	To Be Reviewed

3 Glossary

3.1 Level 1 Data

Level 1 data result from "automatic" pipeline processing of Level 0 data. This processing applies the instrument calibration to remove instrument artifacts and convert the instrument measurements to physical units. If appropriate, Level 1 processing also converts celestial coordinates to J2000, the mission's standard representation. Level 1 data are generally the starting point for scientific analyses.

In LAT Level 1 processing, the Level 0 data, describing the interactions within the LAT, will be analyzed to identify and characterize the interacting particle (e.g., photons, electrons, protons, etc.). Thus tracks will be fitted to the hits in the tracker and calorimeter, the particle trajectories and energies will be estimated, and the event will be classified. The Level 1 data for an event will include not only the parameters from the analysis of the tracks, but also the time, the spacecraft ephemeredes, etc. Other LAT Level 1 data will include histories of the instrument live time and exposure, as well as instrument response functions relevant to the observation.

4 Applicable Documents

- "GLAST Large Area Telescope Flight Investigation: An Astro-Particle Physics Partnership Exploring the High-Energy Universe," P. Michelson, PI.
- GLAST DPWG (Data Products Working Group) Report, Draft 2/15/02, S. Digel http://glast.gsfc.nasa.gov/ssc/Report_DPWG.pdf
- LAT IOC (Instrument Operations Center) System Specification
- GLAST Level I Requirements Document
- Operations Concept Document, 433-OPS-0001
- Science Requirements Document, 433-SRD-0001
- Project Data Management Plan (PDMP), 433-PLAN-0009
- LAT Science Analysis Software Requirements Document, LAT-SS-20.0
- LAT IOC Performance Specification - Level II(B) Specification, LAT-SS-15.1
- LAT Science Analysis Software Management Plan, LAT-MD-360.1
- LAT Event Summary Database Requirements Document Event Database Requirements Draft
- Science Analysis Tools Description Draft

5 Requirements

These requirements relate to the databases used by the GLAST science analysis tools. There are a total of 6 databases identified as D1-D6 by the Science Tools Working Group (Science Analysis Tools Description Draft). The first of these, D1, is the LAT level 1 event summary database and is covered in a separate document (LAT Event Summary Database Requirements Document). D4 is a placeholder for astronomical catalogs external to the GLAST project, and D6 is the Calibration database (CALDB) designed and maintained by the HEASARC. As such, D4 and D6 are outside the scope of this document as we only consider requirements on the databases themselves here.

The three databases covered by this document are: D2 – the pointing, livetime, and mode history database, D3 – the LAT point source catalog, and D5 the pulsar ephemeredes compilation.

The requirements in the next few sections cover the databases themselves. An attempt was made to try to avoid specifying implementation details and concentrate on just what the database has to do.

The first section will introduce the databases. The next section will discuss all the general requirements (with the exception of access related requirements) on the databases . The next section focuses on the access requirements. The final section in the document covers requirements specific to individual databases. The appendices list the expected contents and data sizes for the databases for informational purposes.

5.1 Introduction

The databases covered here are D2, D3, and D5 as defined by the STWG. D2 holds information regarding the spacecraft pointing history, the LAT detector livetime, and the instrument modes selected over the lifetime of the mission. D3 is a catalog of point sources identified by the LAT team as being important for gamma ray science. The final catalog is a compendium of pulsar characteristics that are relevant for gamma ray analyses.

In the next few sections are general requirements that must be met by all three databases. The common elements among them are the main drivers in this requirements document. These are relatively simple databases, so the requirements are not too stringent. It is hoped that a single type of database management system can be used for all databases, which would minimize maintenance complexity for the project.

It is expected that a DataBase Management System (DBMS) will be used to maintain the integrity of the data through ingest and retrieval. However we do not preclude the

possibility of using some system based on flat files (e.g. a Berkeley database system or a custom made management tool) if all the requirements can be met.

5.2 General Requirements

5.2.1 Searchability

Must be able to search on times, integers and reals.

5.2.2 Database handling capacity

Must be able to store and search the entire 10 years worth of mission data.

5.2.3 Operating Systems

The database system must be able to run on hardware and operating systems commonly available in the LAT IOC, the SSC, and the HEASARC.

5.2.4 Maintenance

Must be relatively easy to maintain. Less than 0.2 FTE per database system summed over the three databases here.

5.2.5 Assumption by the HEASARC

The databases shall be maintained by (at least) the HEASARC after the GLAST mission ends

5.2.6 Database system independence

The database system must not require the use of special features to meet these requirements, as this could make changing to a different database system a problem.

5.2.7 Update and backup search concurrency

Must be able to update and backup the database concurrent with searching.

5.2.8 Database Backup

The database must have tools available for incremental and full backup of the data contents.

5.2.9 DBMS specific requirements

These requirements apply if a DBMS system is preferred over some other way to archive the data.

5.2.9.1 SQL version

If a DBMS, must use SQL close to ANSI SQL99 to be compatible with 5.2.5.

5.2.9.2 Primary time key

The databases must be able to use time as a primary search key.

5.3 Access Requirements

5.3.1 Accessibility

Must be accessible through an API in a standard GLAST programming language (e.g., Perl, C++, and Java).

5.3.2 Administrator account

There must be at least one way to restrict write privilege and configuration control to an administrator account.

5.3.3 Remote access

The database must be remotely accessible for queries and data retrieval.

5.3.4 Read Only Web queries

Must be able to restrict some remote database connections to be read only queries.

5.4 Database Specific Requirements

5.4.1 D2. Pointing, livetime, and mode history

5.4.1.1 Functional Description

This is the database of pointing and observation history that is needed to calculate exposures. It contains information about the orientation, location, and operation of the LAT for regular time intervals, ~30 s. The information also includes entries whenever instrument mode changes are made. The analysis tools do not directly access the database. Instead, it receives queries from, and passes data to, the Pointing/livetime history extractor (U2 – see the STWG description).

5.4.1.2 Inputs

Must be able to ingest spacecraft livetime history tables generated by the LIOC for roughly 12 hours periods that are delivered twice per day.

5.4.1.3 Output

The database must be able to deliver a table with all selected rows.

5.4.1.4 Queries required

The database must be able to select rows by time intervals, filtered by other fields in the row.

5.4.1.5 Performance requirements

5.4.1.5.1 Ingest Speed

Must be able to ingest a newly delivered 12-hour data table in < 1 minute.

5.4.1.5.2 Request speed

A standard search is the expected average user query.

5.4.1.5.2.1 Standard Search

The standard search for D2 is to get 6 months worth of consecutive data, which constitutes about 40 Mb of data (TBR).

5.4.1.5.2.2 Service time

Must be able to service a standard search request in < 1minute.

5.4.1.5.2.3 Number of service requests

Must be able to service > 1500 standard search requests in a day.

5.4.1.5.3 Update speed

Must be able to input a re-processed 12-hour data table in < 5 times the time it takes to ingest a brand new table.

5.4.1.5.4 Restore speed

Must be able to rebuild the database from input files in < 1 day.

5.4.2 D3. LAT point source catalog

5.4.2.1 Functional Description

This is the online form of the point source catalog. It is not directly accessed by the analysis tools, but instead receives queries from, and passes data to, the Catalog Access tool (U5 see the STWG description document).

5.4.2.2 Input

Must be able to ingest the latest complete source catalog updated when necessary by LIOC

5.4.2.3 Output

Must be able to output tables with all fields for selected rows.

5.4.2.4 Queries required

Select all entries specified by 2 dimensional region of the sky filtered by selections on other fields (e.g. frequency).

5.4.2.5 Performance requirements

5.4.2.5.1 Ingest Speed

10 Mb of LAT point source data must be able to be ingested and ready for searching in <10 minutes.

5.4.2.5.2 Request speed

A standard search is the expected average user query.

5.4.2.5.2.1 Standard Search

A standard search fetches all sources in an area twice (TBR) the size of a standard event point source search. This is to ensure that there are no strong sources near the edge of the search region that could affect the analysis.

5.4.2.5.2.2 Service time

Must be able to service a standard search in < 1 minute.

5.4.2.5.2.3 Number of service requests

Must be able to satisfy > 1500 service requests per day.

5.4.2.5.3 Update speed

Must be able to update tables of refined point source entries at < 5 times the ingest rate,

5.4.2.5.4 Restore speed

Must be able to recreate the database from the input tables in < 1 day.

5.4.3 D5. Pulsar ephemeredes

5.4.3.1 Functional Description

This is the radio pulsar timing information to be maintained during the LAT mission for assigning pulsar phases to gamma rays. The user does not directly access it, but instead the Pulsar Analysis tool A3 (see the STWG description document) will perform the query. (If the pulsar ephemeredes are implemented in a true database system, then a front-end interface tool, equivalent to the Data Extractor for the gamma-ray data, will be what communicates directly with the database.) The database will also receive queries from, and pass data to, the Catalog Access tool (U5).

5.4.3.2 Inputs

Must be able to ingest new pulsar ephemeredes on LAT team update.
Must also be able to ingest tables with varying numbers of pulsar ephemeredes.

5.4.3.3 Output

Must be able to output tables with all fields for selected rows. Output may be two tables for single and binary pulsars.

5.4.3.4 Queries required

Must be able to select by 2 dimensional region of the sky, filtered by other parameters.
Must be able to select by pulsar period. (TBR).

5.4.3.5 Performance requirements

5.4.3.5.1 Ingest Speed

Must be able to ingest 1 Mb worth of pulsar ephemeredes tables in < 1 minute.

5.4.3.5.2 Request speed

A standard search is the expected average user query.

5.4.3.5.2.1 Standard Search

Standard search is to retrieve all pulsars in a level 1 event database standard search region (a 15 degree radius circle or the equivalent square region.)

5.4.3.5.2.2 Service time

Must be able to perform a standard search of the database in < 1 minute

5.4.3.5.2.3 Number of service requests

Must be able to service >1500 standard requests per day.

5.4.3.5.3 Update speed

Must be able to load an updated database in < 5 times the ingest speed.

5.4.3.5.4 Restore speed

Must be able to restore the database from input data files in < 1 hour.

6 Appendix: Expected Database Characteristics

Listed here are the contents and data sizes anticipated for the databases covered in this document. This information can and will be modified as the Science Tools Working Group continues to refine the detailed specification of these databases.

6.1 D2: Pointing, Livetime, and mode history database

6.1.1 Contents

The provisional contents of the database, defined in the report of the Data Products Working Group as LS-005 and augmented here to include the SAA flag and positions of the sun and moon, are as follows:

	Contents	Units
1	starting time of interval (Mission Elapsed Time)	s
2	ending time of interval (Mission Elapsed Time)	s
3	position of S/C at start of interval (x,y,z inertial coordinates)	km
4	viewing direction at start (LAT +z axis), 2 angles	dimensionless
5	orientation at start (LAT +x axis), 2 angles	dimensionless
6	zenith direction at start, 2 angles	dimensionless
7	LAT operation mode	dimensionless
8	Livetime	s
9	SAA flag	dimensionless
10	S/C longitude	deg
11	S/C longitude	deg
12	S/C altitude	km
13	direction of the sun, 2 angles	Deg
14	direction of the moon, 2 angles	

The positions of the sun and moon are included here solely to facilitate cuts on their positions in the generation of exposure. They are both gamma-ray sources (the sun impulsively) and both of course shadow sources they pass in front of.

6.1.2 Data Sizes

- Table to ingest ~ 100kb
- 1 year data size ~ 73 Mb
- 10 year data size ~730 Mb
- Number of rows in 1 year ~ 63 M rows (30s resolution)

6.2 D3: LAT point Source Catalog

6.2.1 Contents

The provisional contents of the database, defined in the report of the Data Products Working Group as LS-008, are as follows:

	Contents	Units
1	source name (“telephone number”)	dimensionless
2	RA	Deg
3	Dec	Deg
4	th68 semimajor, semiminor axis, and position angle	Deg
5	th95 semimajor, semiminor axis, and position angle	Deg
6	flux (>100 MeV, avg. for the time interval of the catalog)	cm-2 s-1
7	flux uncertainty, 1 sigma (as above)	cm-2 s-1
8	Photon spectral index (avg)	dimensionless
9	variability index	dimensionless
10	significance (avg)	dimensionless
11	significance (peak)	dimensionless
12	peak flux (for time interval above?)	cm-2 s-1
13	peak flux uncertainty	cm-2 s-1
14	time of peak flux (wrt reference date)	S
15	interval of time	S
16	flux history	cm-2 s-1
17	flux uncertainty, 1 sigma (as above)	cm-2 s-1
18	start times of flux history entries	S
19	end times of flux history entries	S
20	candidate counterparts	dimensionless
21	Degrees of confidence for the counterparts	dimensionless
22	flags (confusion, low latitude,…)	dimensionless

6.2.2 Data Size

- Expected catalog size ~ 10 Mb
- Data size after 1 year ~ 10 Mb
- Data size after 10 years ~ 50 Mb (TBR)

6.3 D5: Pulsar Ephemeredes

6.3.1 Contents

The provisional contents of the database, defined based on the pulsar timing files used for EGRET are given below. Note that for generality and consistency with format provided by pulsar observers, times in this file should be specified in MJD rather than Mission Elapsed Time. The second table below contains the additional information required for binary pulsars, but these two tables could be combined.

Parameters for any pulsar

	Contents	Units
1	Pulsar name	dimensionless
2	Right Ascension (J2000)	deg
3	Declination (J2000)	deg
4	Start of interval of validity for timing info (MJD)	days
5	End of interval of validity (MJD)	days
6	Infinite-frequency geocentric UTC arrival time of a pulse (MJD)	days
7	Pulsar rotation frequency	Hz
8	First derivative of pulsar frequency	Hz^2
9	Second derivative of pulsar frequency	Hz^3
10	Root-mean-square radio timing residual (periods)	dimensionless
11	Source of timing information	dimensionless
12	Flag for binary pulsars	dimensionless

Orbital parameters for binary pulsars

	Contents	Units
1	Pulsar name	dimensionless
2	Orbital period	s
3	Projected semi-major axis	s (light travel time)
4	Orbital eccentricity	dimensionless
5	Barycentric time (TDB scale) of periastron (MJD)	days
6	Longitude of periastron	deg
7	First derivative of longitude of periastron	deg per Julian year
8	Time-dilation and gravitational redshift parameter	s
9	First derivative of orbital period	dimensionless
10	Source of orbital parameters	dimensionless

6.3.2 Data Sizes

- Expected catalog size ~ 1 Mb (1000 pulsars entries of 1 kb each)
- Data size after 1 year ~ 1 Mb
- Data size after 10 years ~ 5 Mb (TBR)